Supplemental Material for "Three-dimensional color code thresholds via statistical-mechanical mapping"

Aleksander Kubica,[1,2,3] Michael E. Beverland,[1,4] Fernando Brandão,[4,1] John Preskill,[1] and Krysta M. Svore[4]

[1] *Institute for Quantum Information & Matter, California Institute of Technology, Pasadena, CA 91125, USA*
[2] *Perimeter Institute for Theoretical Physics, Waterloo, ON, N2L 2Y5, Canada*
[3] *Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1, Canada*
[4] *Station Q Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA*

HEURISTIC ESTIMATES OF THE THRESHOLD

For a family of topological CSS codes, such as the toric or color codes, one can use the following, *nonrigorous* reasoning to arrive at the *heuristic estimates* of thresholds p^X and p^Z for the bit-flip and phase-flip noise models with optimal decoders and perfect measurements. Let L denote the linear system size, $n = n(L)$ be the number of physical qubits, $N_X = N_X(L)$ be the number of independent X-type stabilizers, and p be the phase-flip error rate. Thus, the average number of Z-type errors in the system is np, the number of different error patterns is $\binom{n}{np}$, and the number of different X-type syndromes is 2^{N_X}. As long as the error rate $p \leq p^Z$ we expect the following inequality to hold

$$\text{number of } Z\text{-error configurations} \lesssim \text{number of } X\text{-syndrome configurations}, \tag{A.1}$$

which after taking the logarithm of both sides and using Stirling's approximation leads to $nH(p) \simeq \log_2 \binom{n}{np} \lesssim N_X$, where $H(p) = -p \log_2 p + (1-p)\log_2(1-p)$ is the Shannon entropy. Thus, the heuristic estimate of the phase-flip threshold p^Z is obtained as the solution of the equation

$$H(p^Z) = \lim_{L\to\infty} \frac{N_X}{n}, \tag{A.2}$$

such that $p^Z \leq 1/2$; one similarly obtains the estimate of the bit-flip threshold p^X.

Let us consider the 3D color code on the bcc lattice of linear size L, which contains $V = 2L^3$ vertices, $E = 14L^3$ edges, $F = 24L^3$ faces and $T = 12L^3$ tetrahedra. Since the number of physical qubits is $n = T$ and the number of independent X-stabilizers is $N_X = V - 3$, thus by solving Eq. (A.2) we arrive at the heuristic estimate of the threshold $p^Z_{\text{3DCC}} \simeq 0.0246$. Similarly, we obtain the heuristic estimate of the threshold $p^X_{\text{3DCC}} \simeq 0.2644$, where we use the fact that the number of independent Z-stabilizers is $N_Z = E - 2V - 6$, see [13]. We emphasize that the threshold estimates p^Z_{3DCC} and p^X_{3DCC} are not rigorous and should serve only as an intuition about the actual values $p^{(1)}_{\text{3DCC}} \simeq 1.9\%$ and $p^{(2)}_{\text{3DCC}} \simeq 27.6\%$. We remark that by applying the same reasoning to the 3D toric code on the cubic lattice we find the thresholds for the phase- and bit-flip noise models to be $p^Z_{\text{3DTC}} \simeq 0.0615$ and $p^X_{\text{3DTC}} \simeq 0.1740$, which should be compared with the actual values $p^{(1)}_{\text{3DTC}} \simeq 3.3\%$ and $p^{(2)}_{\text{3DTC}} \simeq 23.5\%$.

DUALITY OF MODELS FOR ZERO DISORDER

We already mentioned that the 4- and 6-body RCIM described by Eqs. (11) and (12) are dual for $p = 0$, i.e., the case with no disorder. Here we say that two models are dual if the low-temperature expansion of the partition function of one model matches the high-temperature expansion of the partition function of the other and vice versa. For concreteness, we consider the spin models on the bcc lattice, however similar reasoning is applicable to any valid color code lattice. Let us follow the standard procedure [3, 35] and write a series expansion for the partition function of the 4-body RCIM in two limits.

- Low temperature $\beta^X \gg 1$ — there are 8 ground states, which are related by the global $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry to the ferromagnetic state with $s_v = +1$ for all spins and thus one can expand the partition function by including low energy excitations around the ground space to obtain

$$Z^X_0(\beta^X) = \sum_{\{s_v = \pm 1\}} e^{-\beta^X H^X_0(\{s_v\})} \propto e^{\beta^X |\Delta_3(\mathcal{L})|}(1 + |\Delta_0(\mathcal{L})| e^{-48\beta^X} + \ldots). \tag{A.3}$$

- High temperature $\beta^X \ll 1$ — we treat spins independently and expand the partition function in powers of $\tanh\beta^X$ to obtain

$$Z_0^X(\beta^X) = \sum_{\{s_v=\pm1\}} e^{-\beta H_0^X(\{s_v\})} = \sum_{\{s_v=\pm1\}} \prod_{t\in\Delta_3(\mathcal{L})} \exp\left(\beta^X \right) \tag{A.4}$$

$$= (\cosh\beta^X)^{|\Delta_3(\mathcal{L})|} \sum_{\{s_v=\pm1\}} \prod_{t\in\Delta_3(\mathcal{L})} \left(1 + \tanh\beta^X\right) \tag{A.5}$$

$$= 2^{|\Delta_0(\mathcal{L})|}(\cosh\beta^X)^{|\Delta_3(\mathcal{L})|}(1 + E_4(\tanh\beta^X)^4 + E_6(\tanh\beta^X)^6 + \ldots), \tag{A.6}$$

  where $E_4 = \frac{3}{7}\Delta_1(\mathcal{L})$ and $E_6 = \frac{4}{7}\Delta_1(\mathcal{L})$ are the numbers of edges whose neighborhoods (i.e., 3-stars) contain 4 and 6 tetrahedra, respectively[1]. Note that contributions to the second and third terms in the expansion are due to selecting an edge $e \in \Delta_1(\mathcal{L})$ and multiplying all 4 or 6 (depending on the edge) weight-four spin operators identified with tetrahedra containing e, which results in all spin terms canceling each other.

In a similar way we can find a series expansion for the partition function of the 6-body RCIM on the bcc lattice in the limit of low and high temperatures.

- Low temperature $\beta^Z \gg 1$ — the number of lowest-energy configurations (related by local $\mathbb{Z}_2 \times \mathbb{Z}_2$ gauge transformations) is proportional to $2^{2|\Delta_0(\mathcal{L})|}$ and expanding around the ground space

$$Z_0^Z(\beta^Z) = \sum_{\{s_e=\pm1\}} e^{-\beta^Z H_0^Z(\{s_e\})} \propto 2^{2|\Delta_0(\mathcal{L})|} e^{\beta^Z|\Delta_3(\mathcal{L})|}(1 + E_4 e^{-8\beta^Z} + E_6 e^{-12\beta^Z} + \ldots). \tag{A.7}$$

- High temperature $\beta^Z \ll 1$ — expand the partition function in powers of $\tanh\beta^Z$ to obtain

$$Z_0^Z(\beta^Z) = (\cosh\beta^Z)^{|\Delta_3(\mathcal{L})|} \sum_{\{s_e=\pm1\}} \prod_{t\in\Delta_3(\mathcal{L})} \left(1 + \tanh\beta^Z\right) \tag{A.8}$$

$$= 2^{|\Delta_1(\mathcal{L})|}(\cosh\beta^Z)^{|\Delta_3(\mathcal{L})|}(1 + |\Delta_0(\mathcal{L})|(\tanh\beta^Z)^{24} + \ldots), \tag{A.9}$$

  where contributions to the second term in the expansion are due to selecting a vertex $v \in \Delta_0(\mathcal{L})$ and multiplying all 24 weight-six spin operators identified with tetrahedra in the neighborhood (i.e., the 3-star) of v so that spin terms cancel each other.

We conclude that if $e^{-2\beta^Z} = \tanh\beta^X$ (or equivalently $e^{-2\beta^X} = \tanh\beta^Z$), then the terms in the series expansions in Eqn. (A.6) and (A.7) match each other up to a constant prefactor; similarly for terms in Eqn. (A.3) and (A.9). In particular, if there is only one phase transition in the first model at temperature T_c^X, then there is a unique phase transition in the dual model at temperature

$$T_c^Z = -2\left(\log\tanh\frac{1}{T_c^X}\right)^{-1}. \tag{A.10}$$

This serves as a consistency check for our results. Indeed, for zero disorder $p = 0$ the critical temperatures $T_c^X = 8.77(2)$ and $T_c^Z = 0.918(3)$ for the 4- and 6-body RCIM are related according to Eq. (A.10) within the statistical uncertainty.

PROOF OF IMPLICATION

Here we show that successful decoding implies diverging average energy cost of introducing any non-trivial domain wall. We used this fact in the derivation of statistical-mechanical models to relate the threshold p_c of optimal error correction to the critical point p_N on the Nishimori line. Note that this implication allows us to only infer that $p_c \leq p_N$. However, we expect that successful decoding be possible throughout the ordered phase and thus these two values should coincide.

---

[1] If we choose the coordinates of vertices of the bcc lattice to be triples of either integers or half-integers, then edges surrounded by 4 tetrahedra are aligned with directions $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, whereas edges surrounded by 6 tetrahedra are aligned with directions $(\pm1,\pm1,\pm1)$.

Lemma 1. *Consider a CSS code described by the chain complex in Eq. (3). Let $H_1 = \ker \partial_1 / \operatorname{im} \partial_2$ be the first homology group of finite cardinality, $|H_1| < \infty$. If the probability of successful optimal X-error correction goes to 1 in the limit of infinite system size*

$$\mathrm{pr}(\mathrm{succ}) = \sum_{\epsilon \in C_1} \mathrm{pr}(\epsilon)\mathrm{pr}(\mathrm{succ}|\epsilon) \to 1, \tag{A.11}$$

then the average free energy cost of introducing any non-trivial domain wall $\lambda \in \ker \partial_1 \setminus \operatorname{im} \partial_2$ diverges

$$\langle \Delta_\lambda \rangle = \sum_{\epsilon \in C_1} \mathrm{pr}(\epsilon)\Delta_\lambda(\epsilon) \to \infty. \tag{A.12}$$

Proof. Let $\overline{\epsilon} = \{\epsilon + \partial_2\omega | \omega \in C_2\}$ denote the equivalence class of errors for $\epsilon \in C_1$ and $\mathcal{E} = \{\overline{\epsilon}, \ldots\}$ be the set of all equivalence classes. We define a representative of the most probable equivalence class of errors consistent with the syndrome $\sigma \in C_0$ to be

$$\rho(\sigma) = \arg \max_{\substack{\rho \in C_1 \\ \partial_1 \rho = \sigma}} \mathrm{pr}(\overline{\rho}). \tag{A.13}$$

The conditional probability of successful decoding using the optimal (maximum likelihood) decoder is given by

$$\mathrm{pr}(\mathrm{succ}|\epsilon) = \begin{cases} 1 \text{ if } \epsilon \in \overline{\rho(\partial_1 \epsilon)}, \\ 0 \text{ otherwise.} \end{cases} \tag{A.14}$$

Thus, we have

$$\mathrm{pr}(\mathrm{succ}) = \sum_{\epsilon \in C_1} \mathrm{pr}(\epsilon)\mathrm{pr}(\mathrm{succ}|\epsilon) = \sum_{\sigma \in \operatorname{im} \partial_1} \mathrm{pr}(\overline{\rho(\sigma)}). \tag{A.15}$$

By rewriting the sum over all equivalence classes of errors $\overline{\epsilon} \in \mathcal{E}$ as the sum over all possible syndromes $\sigma \in \operatorname{im} \partial_1$ and different representatives $\lambda' \in H_1$ of the homology group we arrive at

$$1 = \sum_{\overline{\epsilon} \in \mathcal{E}} \mathrm{pr}(\overline{\epsilon}) = \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'}) = \mathrm{pr}(\mathrm{succ}) + \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{0 \neq \lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'}). \tag{A.16}$$

We want to show two inequalities

$$\mathrm{pr}(\mathrm{succ}) \geq \sum_{\overline{\epsilon} \in \mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\epsilon + \lambda'})} \geq 2\mathrm{pr}(\mathrm{succ}) - 1. \tag{A.17}$$

In order to show the first inequality (A.17) note that

$$\mathrm{pr}(\mathrm{succ}) = \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{\lambda'' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda''}) \frac{\mathrm{pr}(\overline{\rho(\sigma)})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})} \tag{A.18}$$

$$\geq \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{\lambda'' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda''}) \frac{\mathrm{pr}(\overline{\rho(\sigma) + \lambda''})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})} = \sum_{\overline{\epsilon} \in \mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\epsilon + \lambda'})}, \tag{A.19}$$

where we use $\mathrm{pr}(\overline{\rho(\sigma)}) \geq \mathrm{pr}(\overline{\rho(\sigma) + \lambda''})$ for all $\sigma \in \operatorname{im} \partial_1$ and $\lambda'' \in H_1$. The second inequality (A.17) follows from

$$\mathrm{pr}(\mathrm{succ}) = \sum_{\sigma \in \operatorname{im} \partial_1} \mathrm{pr}(\overline{\rho(\sigma)}) = \sum_{\sigma \in \operatorname{im} \partial_1} \mathrm{pr}(\overline{\rho(\sigma)}) \frac{\mathrm{pr}(\overline{\rho(\sigma)})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})} + \sum_{\sigma \in \operatorname{im} \partial_1} \mathrm{pr}(\overline{\rho(\sigma)}) \frac{\sum_{0 \neq \lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})} \tag{A.20}$$

$$\leq \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{\lambda'' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda''}) \frac{\mathrm{pr}(\overline{\rho(\sigma) + \lambda''})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'})} + \sum_{\sigma \in \operatorname{im} \partial_1} \sum_{0 \neq \lambda' \in H_1} \mathrm{pr}(\overline{\rho(\sigma) + \lambda'}) \tag{A.21}$$

$$= \sum_{\overline{\epsilon} \in \mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda' \in H_1} \mathrm{pr}(\overline{\epsilon + \lambda'})} + (1 - \mathrm{pr}(\mathrm{succ})). \tag{A.22}$$

If $\mathrm{pr}(\mathrm{succ}) \to 1$, then from inequalities (A.17) we infer that

$$\sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} \to 1, \tag{A.23}$$

and thus for $\lambda \in \ker \partial_1 \setminus \mathrm{im}\, \partial_2$ we have

$$\sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon+\lambda})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} \to 0. \tag{A.24}$$

In the last step we used the following inequalities

$$0 \leq \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon+\lambda})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} \leq \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\sum_{0\neq\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} = 1 - \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})}. \tag{A.25}$$

We rewrite $\langle \Delta_\lambda \rangle$ in the following way

$$\langle \Delta_\lambda \rangle = \sum_{\epsilon\in C_1} \mathrm{pr}(\epsilon)\Delta_\lambda(\epsilon) = \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \log \frac{\mathrm{pr}(\overline{\epsilon})}{\mathrm{pr}(\overline{\epsilon+\lambda})} \tag{A.26}$$

$$= \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \log \frac{\mathrm{pr}(\overline{\epsilon})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} - \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \log \frac{\mathrm{pr}(\overline{\epsilon+\lambda})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})}. \tag{A.27}$$

Using the inequality $\log x \geq 1 - \frac{1}{x}$ to lower-bound the first term and Jensen inequality for the second term we obtain

$$\langle \Delta_\lambda \rangle \geq (1 - |H_1|) - \log \sum_{\overline{\epsilon}\in\mathcal{E}} \mathrm{pr}(\overline{\epsilon}) \frac{\mathrm{pr}(\overline{\epsilon+\lambda})}{\sum_{\lambda'\in H_1} \mathrm{pr}(\overline{\epsilon+\lambda'})} \to \infty. \tag{A.28}$$

□

FINDING PHASE TRANSITIONS

In order to map the disorder-temperature phase diagrams of the 4- and 6-body RCIM in Fig. 1 we need to reliably identify phase transitions. Here we describe in detail how we achieve that by analyzing specific heat, the spin-spin correlation function and the Wilson loop operator. We also present additional results for the 4- and 6-body RCIM in Fig. A.1.

A. Specific heat

For a second-order phase transition, the specific heat $c(T)$ as a function of temperature T is expected to have a discontinuity near a phase transition at temperature T_c in the limit of infinite system size $L \to \infty$. However, for a system of finite linear size L, the peak of the specific heat $c_L(T)$ appears at temperature $T_c(L) = \arg\max_T c_L(T)$ shifted from that in the infinite system by an amount

$$\left| \frac{T_L - T_c}{T_c} \right| \propto L^{-1/\nu}, \tag{A.29}$$

where ν is the correlation length critical exponent [39]. A similar scaling behavior has been established for first-order phase transitions [5–8]. Thus, we find the critical temperature T_c by fitting a function

$$T_c(L) \sim aL^{-b} + T_c \tag{A.30}$$

to the position of the specific heat peaks for different system sizes and evaluating $T_c(L = \infty)$.

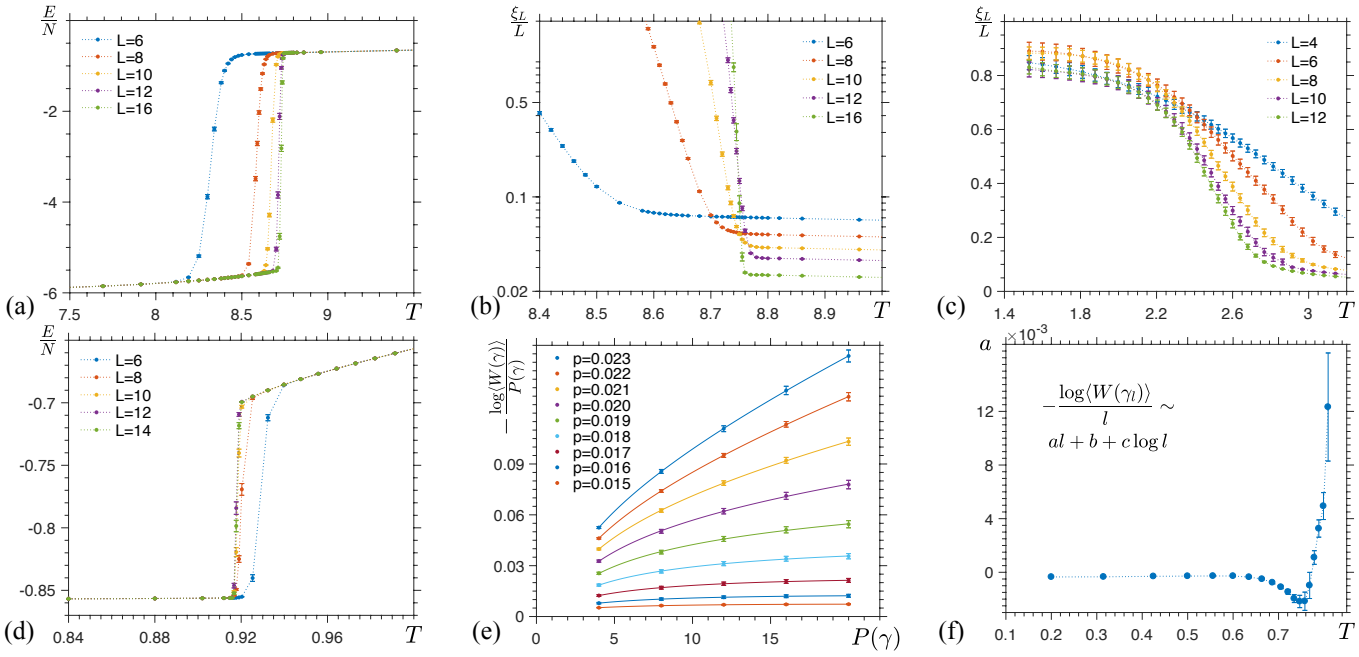

FIG. A.1. Additional details about the 4-body (a)-(c) and 6-body (d)-(f) RCIM. The discontinuity in energy per spin E/N in (a) and (d) suggests first-order phase transitions for both models for $p = 0$. (b) For $p = 0$ the normalized correlation length ξ_L/L does not seem to be described well by the scaling ansatz in Eq. (15) possibly due to a transition being first-order. (c) For the disorder value $p = 0.276$ close to the critical point on the Nishimori line $p_N = p^{(2)}_{\text{3DCC}}$ detecting a phase transition and estimating its critical temperature becomes difficult. (e) We check if the Wilson loop operator $W(\gamma)$ satisfies the perimeter law by plotting $-\log\langle W(\gamma)\rangle/P(\gamma)$ as a function of perimeter $P(\gamma)$ of the square loop γ for different disorder values p, fixed temperature $T = 0.42$ and $L = 10$. We see a change of scaling as the system undergoes a phase transition at $p = 0.019(1)$. (f) We find a fit $-\log\langle W(\gamma_l)\rangle/l \sim al + b + c\log l$ to the data for $p = 0.018$ in Fig. 3(c) and plot the fit parameter a as a function of temperature T. We identify the critical temperature $T_c = 0.75(3)$ of a transition as a location where $a = 0$.

B. Correlation function

One might not be able to identify a phase transition of higher order by looking at the specific heat. Rather, one needs to analyze the behavior of e.g. the order parameter correlation length ξ. In particular, for the system of finite size L and with fixed disorder strength p we define the two-point finite-size correlation length ξ_L as a function of temperature T

$$\xi_L(T) = \frac{1}{2\sin(k_0/2)}\sqrt{\frac{\langle\chi(\vec{0})\rangle}{\langle\chi(\vec{k}_0)\rangle} - 1}, \tag{A.31}$$

where $\langle\chi(\vec{k})\rangle = \sum_{\epsilon\subset\Delta_3(\mathcal{L})} \text{pr}(\epsilon)\chi(\vec{k})$, $\vec{k}$ is the wavevector and $\vec{k}_0 = (2\pi/L, 0, 0)$. In above, we use the thermal expectation value of the wavevector-dependent sublattice magnetic susceptibility

$$\chi(\vec{k}) = \sum_{\{s_v\}} \frac{1}{N}\left(\sum_{u\in U} s_u e^{i\vec{k}\cdot\vec{r_u}}\right)^2 \frac{e^{-\beta H^X_\epsilon(\{s_v\})}}{Z_\epsilon(\beta)}. \tag{A.32}$$

where $\vec{r_u}$ denotes the position of the vertex spin s_u in a sublattice $U \subset \Delta_0(\mathcal{L})$ of single-color vertices. Near a phase transition at temperature T_c, the normalized correlation length is expected to scale as

$$\frac{\xi_L(T)}{L} \sim f(L^{1/\nu}(T - T_c)), \tag{A.33}$$

where f is a dimensionless scaling function and ν is the correlation length critical exponent. We can estimate T_c by plotting $\xi_L(T)/L$ as a function of temperature T for different system sizes L and finding their crossing point. If there is no crossing, then we conclude that there is no phase transition.

C. Wilson loop operator

When the system under consideration has a local (gauge) symmetry, one cannot use a local order parameter to detect a phase transition. Rather, one needs to consider gauge-invariant quantities, such as the Wilson loop operator $W(\gamma)$ in Eq. (16). Suppose γ is a square loop. We denote by $P(\gamma)$ and $A(\gamma)$ the perimeter of γ and the minimal area enclosed by γ, respectively. The scaling of the averaged Wilson loop operator $\langle W(\gamma)\rangle$ in the limit of large loops changes between the ordered (Higgs) and disordered (confinement) phases. Namely,

- in the disordered phase: $\langle W(\gamma)\rangle \sim \exp(-\,\mathrm{const}\cdot A(\gamma))$,
- in the ordered phase: $\langle W(\gamma)\rangle \sim \exp(-\,\mathrm{const}\cdot P(\gamma))$.

We consider a system of finite size L and denote by γ_l a square loop of linear size $l \leq L/2$. Since $A(\gamma_l) \propto l^2$ and $P(\gamma_l) \propto l$, then $\log\langle W(\gamma_l)\rangle$ should scale either quadratically or linearly in l, depending on the phase of the system. Due to finite-size effects, there are some corrections to the area and perimeter scaling. In particular, we numerically find that

$$-\frac{\log\langle W(\gamma_l)\rangle}{l} \sim al + b + c\log l, \tag{A.34}$$

where a, b, c are fitting parameters. We identify the disordered phase as the region where the fitting parameter a is positive, $a > 0$. We use the ansatz in Eq. (A.34) to fit the data in Figs. 3(e)(f), A.1(e) and A.2(b)-(e).

ISING LATTICE GAUGE THEORY

As an example of using specific heat and the scaling of the Wilson loop operator to identify a phase transition we study a known model, the three-dimensional random plaquette Ising model (RPIM); see Fig. A.2. The RPIM is a generalization of $\mathbb{Z}_2$ Ising lattice gauge theory [35], which is relevant for studying the optimal error correction threshold for 1D string-like operators in the 3D toric code [3]. The RPIM is a statistical-mechanical model with classical spins $s_e = \pm 1$ placed on edges $e \in \Delta_1(\mathcal{C})$ of the cubic lattice $\mathcal{C}$ and disorder $\epsilon \subset \Delta_2(\mathcal{C})$. The Hamiltonian describing the RPIM

$$H_\epsilon^{\mathrm{RPIM}}(\{s_e\}) = -\sum_{f\in\Delta_2(\mathcal{C})} (-1)^{[\epsilon]_f}$$

(A.35)

contains 4-body terms, which are products of four edge spins around every square face $f \in \Delta_2(\mathcal{C})$ of the lattice $\mathcal{C}$. We set $[\epsilon]_f = 1$ if $f \in \epsilon$, otherwise $[\epsilon]_f = 0$. We observe that $H_\epsilon^{\mathrm{RPIM}}(\{s_e\})$ has a local $\mathbb{Z}_2$ symmetry, generated by flips of spins on all edges incident on any vertex $v \in \Delta_0(\mathcal{C})$. The Wilson loop operator $W(\gamma_l)$ is a gauge-invariant quantity, where γ_l is a square loop of linear size l. The disorder-temperature phase diagram of the 3D RPIM is shown in Fig. A.3.

NUMERICAL SIMULATION DETAILS

The numerical complexity of simulating the statistical-mechanical models, such as the 4- and 6-body RCIM and the RPIM, increases with the disorder strength p, which is reminiscent of a spin glass behavior. To speed up simulations we use the parallel tempering technique. The parallel tempering technique requires simultaneous simulation of multiple copies $k = 1, \ldots, N_T$ of the system with the same disorder ϵ but different spin configurations $\{s_i\}_k$ and temperatures $T_1 < \ldots < T_{N_T}$. After performing single-spin Metropolis updates for all spins in every copy of the system, swaps of spin configurations $\{s_i\}_k \leftrightarrow \{s_i\}_{k+1}$ of copies at neighboring temperatures T_k and T_{k+1} are allowed with probability

$$\mathrm{pr}(k \leftrightarrow k+1) = \exp\left((E_k - E_{k+1})\left(\frac{1}{T_k} - \frac{1}{T_{k+1}}\right)\right), \tag{A.36}$$

where E_k and E_{k+1} denote energies of spin configurations $\{s_i\}_k$ and $\{s_i\}_{k+1}$. We choose temperatures $T_1 < \ldots < T_{N_T}$ is such a way that the exchange rate $\{s_i\}_k \leftrightarrow \{s_i\}_{k+1}$ estimated by counting the number of (successful) swaps of systems at neighboring temperatures is constant up to statistical fluctuations; for more in-depth discussions see e.g. [11]. Equilibration of the system is tested by a logarithmic binning of data. Namely, the total number of time steps of the simulation is $1 + t_1 + t_2 + \ldots + t_\tau = 2^\tau$, where $t_i = 2^{i-1}$ and one time step consists of an update of every

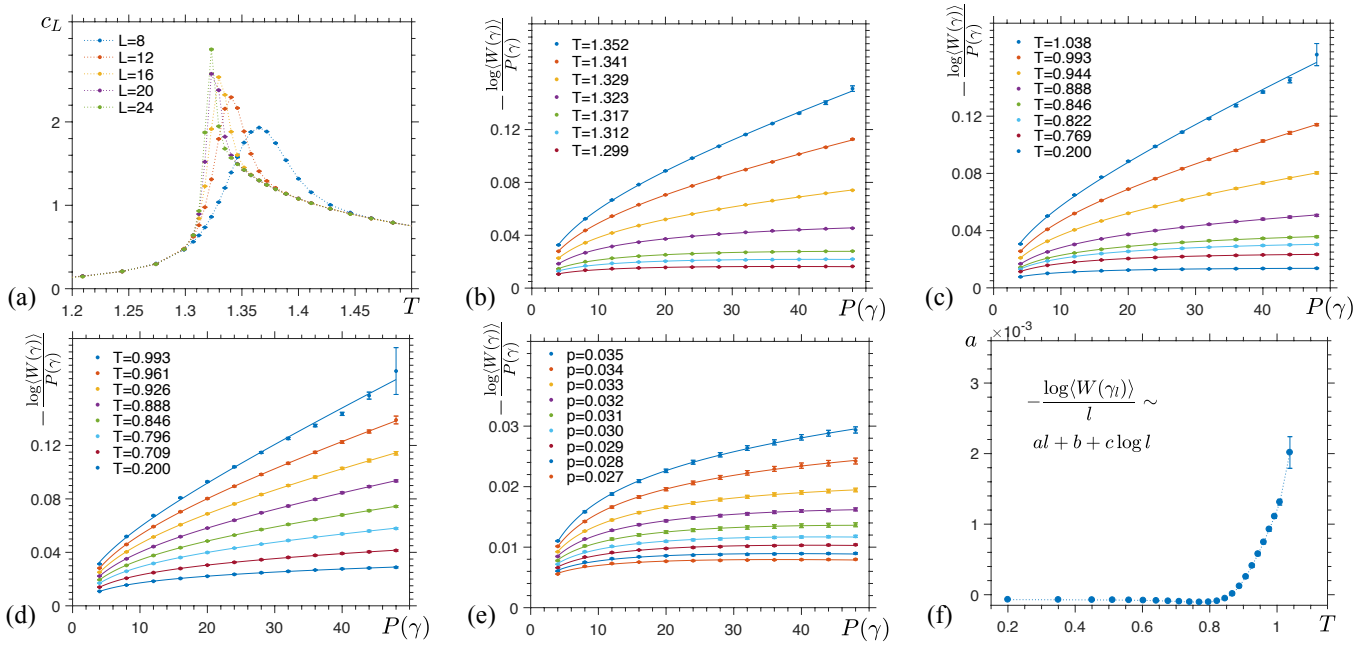

FIG. A.2. Results for the 3D RPIM. (a) For $p = 0$ we can estimate the critical temperature $T_c = 1.316(4)$ of a phase transition by finding the peak positions of specific heat c_L for different system sizes L and exploiting finite-size scaling. In (b)-(d) we check for $p = 0$, $p = 0.031$ and $p = 0.035$ whether the Wilson loop operator $W(\gamma)$ satisfies the perimeter law by plotting $-\log\langle W(\gamma)\rangle/P(\gamma)$ as a function of perimeter $P(\gamma)$ of the square loop γ for different temperatures T and $L = 24$. (e) For fixed temperature $T = 0.45$ we analyze scaling of $-\log\langle W(\gamma)\rangle/P(\gamma)$ for different disorder values p. (f) We find a fit $-\log\langle W(\gamma_l)\rangle/l \sim al + b + c\log l$ to the data in (c) and plot the fit coefficient a as a function of temperature T. We identify the critical temperature $T_c = 0.84(3)$ of a phase transition in (c) as a location where $a = 0$. In (b),(c) and (e) we see a change of scaling as the system undergoes a phase transition at $T_c = 1.317(6)$, $T_c = 0.84(3)$ and $p_c = 0.032(1)$, respectively. In (d) there is no indication of a transition.

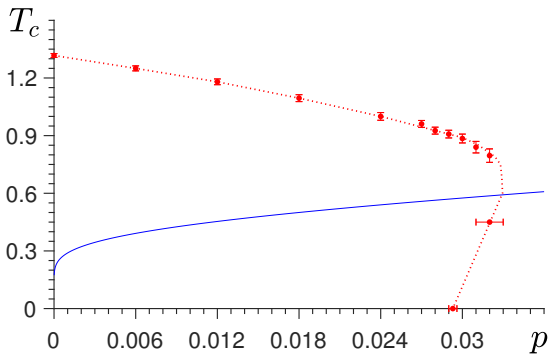

FIG. A.3. The disorder-temperature (p, T)-phase diagram of the 3D random plaquette Ising model on the cubic lattice. The intersection of the Nishimori line (blue) with the anticipated phase boundary (red dotted line) gives the 3D toric code threshold $p^{(1)}_{\text{3DTC}} \simeq 3.3\%$ for optimal error correction associated with 1D string-like logical operators (and point-like excitations). Note that the location of a phase transition for $T = 0$ was found in [23].

spin in all N_T copies of the system followed by (attempted) swaps $\{s_i\}_k \leftrightarrow \{s_i\}_{k+1}$ of spin configurations. We say that the system equilibrates if the quantities of interest, such as the correlation length or the Wilson loop operator, evaluated based on the data from the last three periods of time $t_{\tau-2}$, $t_{\tau-1}$ and t_τ agree up to statistical uncertainty, see e.g. [26, 28]. Numerical simulation details for the 4-body RCIM, the 6-body RCIM and the RPIM are presented in Table A.1.

To estimate statistical error bars of quantities analyzed in the simulation we use the bootstrap technique. The main idea behind the bootstrap technique is to repeat sampling from the existing data set $D = \{d_1, \ldots, d_m\}$ and evaluating a quantity of interest $q = q(D)$. In particular, for $i = 1, \ldots, n$ we perform the following steps

1. from the data set D randomly choose m data points $d_{i(j)}$, where $i(j) \in \{1, \ldots, m\}$,
2. evaluate the quantity $q_i = q(D_i)$ from the data set $D_i = \{d_{i(1)}, \ldots, d_{i(m)}\}$.

Note that in step 1 we allow to choose the same data point multiple times. The relevant quantity q is estimated to be

$$q = \bar{q} \pm \sqrt{\sum_{i=1}^{n} \frac{(\bar{q} - q_i)^2}{n-1}}, \tag{A.37}$$

where $\bar{q} = \frac{1}{n} \sum_{i=1}^{n} q_i$.

---

[13] A. Kubica, B. Yoshida, and F. Pastawski, New Journal of Physics **17**, 083026 (2015).
[35] F. J. Wegner, Journal of Mathematical Physics **12**, 2259 (1971).
[3] M. Kardar, *Statistical physics of fields* (Cambridge University Press, 2007).
[39] N. Goldenfeld, *Lectures on phase transitions and critical phenomena* (Addison-Wesley Publishing Company, 1992).
[5] M. S. S. Challa, D. P. Landau, and K. Binder, Physical Review B **34**, 1841 (1986).
[6] K. Binder, Rep. Prog. Phys. **50**, 783 (1987).
[7] J. Lee and J. M. Kosterlitz, Physical Review B **43**, 3265 (1991).
[8] C. Borgs and R. Kotecky, Physical Review Letters **68**, 1734 (1992).
[3] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Journal of Mathematical Physics **43**, 4452 (2002).
[23] C. Wang, J. Harrington, and J. Preskill, Annals of Physics **303**, 31 (2003).
[11] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, Journal of Statistical Mechanics: Theory and Experiment **2006**, 12 (2006).
[26] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, Physical Review Letters **103**, 090501 (2009).
[28] R. S. Andrist, *Understanding topological quantum error-correction codes using classical spin models*, Ph.D. thesis (2012).

p	$L_{\max}$	N_ϵ	τ	N_T	$T_{\min}$	$T_{\max}$
0.000	16	500	20	55	2.40	12.80
0.050	16	500	20	42	2.30	11.40
0.100	16	500	20	41	2.20	10.15
0.150	16	500	20	42	2.10	8.42
0.200	16	500	20	41	2.00	6.80
0.250	16	500	20	42	1.90	4.97
0.265	16	500	20	34	1.80	3.53
0.270	16	500	20	34	1.60	3.32
0.272	12	500	20	34	1.60	3.30
0.274	12	500	20	34	1.53	3.21
0.276	12	500	20	34	1.53	3.21
0.280	12	500	20	34	1.33	3.18
0.000	14	250	20	47	0.20	1.28
0.003	12	250	20	44	0.20	1.25
0.006	12	250	20	42	0.20	1.22
0.009	12	250	20	39	0.20	1.17
0.012	12	250	20	38	0.20	1.14
0.015	10	250	22	37	0.20	1.08
0.016	10	250	22	37	0.20	1.08
0.017	10	250	22	37	0.20	1.08
0.018	10	250	22	37	0.20	1.08
0.019	10	250	22	37	0.20	1.08
0.020	10	250	22	37	0.20	1.08
0.021	10	250	22	37	0.20	1.08
0.000	24	500	19	51	0.40	2.08
0.006	24	500	19	43	0.40	1.95
0.012	24	500	19	41	0.40	1.77
0.018	24	500	19	43	0.35	1.64
0.024	24	500	19	42	0.30	1.49
0.027	24	250	19	43	0.20	1.28
0.028	24	250	19	43	0.20	1.28
0.029	24	250	19	43	0.20	1.28
0.030	24	250	19	43	0.20	1.28
0.031	24	250	19	43	0.20	1.28
0.032	24	250	19	43	0.20	1.28
0.023	24	250	19	43	0.20	1.28
0.034	24	250	19	43	0.20	1.28
0.035	24	250	19	43	0.20	1.28

TABLE A.1. Numerical simulation parameters for: the 4-body RCIM (top), the 6-body RCIM (middle), and the IGT (bottom). $L_{\max}$ and N_ϵ denote the linear size of the biggest simulated system and the number of randomly chosen disorder samples. N_T denotes the number of temperatures in the range $[T_{\min}, T_{\max}]$ chosen in a way that the exchange rate of spin configurations is approximately constant. 2^τ is the total number of time steps of the simulation, where one time step consists of an update of every spin in all N_T copies of the system followed by (attempted) swaps $\{s_i\}_k \leftrightarrow \{s_i\}_{k+1}$ of spin configurations.